DOCUMENT RESUME

ED 463 299 TM 033 726

AUTHOR	Harwell, Michael; Serlin, Ronald C.
TITLE	An Empirical Study of Eight Nonparametric Tests in Hierarchical Regression.
PUB DATE	2002-04-11
NOTE	39p.; Paper prepared for the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Chi Square; Computer Simulation; Data Analysis; Mathematical Models; *Nonparametric Statistics; *Regression (Statistics)
IDENTIFIERS	Confidence Intervals (Statistics); *F Test; *Hierarchical Analysis

ABSTRACT

When normality does not hold, nonparametric tests represent an important data-analytic alternative to parametric tests. However, the use of nonparametric tests in educational research has been limited by the absence of easily performed tests for complex experimental designs and analyses, such as factorial designs and multiple regression analyses, and limited information about the properties of these tests for realistic data conditions. Efforts to remedy this deficiency have begun with the introduction of general linear model-based nonparametric tests. The results of a computer simulation of the properties of several of these tests in hierarchical regression analyses indicated that, on balance, the top-performing test for the nonnormal distributions studied was the McKean-Hettmansperger F-test (J. McKean and T. Hettmansperger, 1976) using a confidence interval estimate of the scale parameter tau. The Serlin-Harwell aligned-rank chi-square test (R. Serlin and M. Harwell, 2001) performed almost as well, which, combined with the fact that it is easier to compute, makes it an attractive competitor to the Mc-Kean-Hettmansperger test. (Contains 5 tables and 59 references.) (Author/SLD)

An Empirical Study of Eight Nonparametric Tests in Hierarchical Regression

Michael Harwell
University of Minnesota

Ronald C. Serlin
University of Wisconsin-Madison

March 11, 2002

Paper to be presented at the annual meeting of the American Educational Research Association, New Orleans, April.
Correspondence concerning this paper should be directed to Michael Harwell, Department of Educational Psychology, 178 Pillsbury Drive, SE, University of Minnesota, Minneapolis, MN 53455 [email: harwe001@umn.edu]

Abstract

When normality does not hold, nonparametric tests represent an important data-analytic alternative to parametric tests. However, the use of nonparametric tests in educational research has been limited by the absence of easily performed tests for complex experimental designs and analyses, such as factorial designs and multiple regression analyses, and limited information about the properties of these tests for realistic data conditions. Efforts to remedy this deficiency have begun with the introduction of general linear model-based nonparametric tests. The results of a computer simulation of the properties of several of these tests in hierarchial regression analysis indicated that, on balance, the top-performing test for the nonnormal distributions studied was the McKean-Hettmansperger F-test using a confidence interval estimate of the scale parameter τ. The Serlin-Harwell aligned-rank chi-square test performed almost as well, which, combined with the fact that it is easier to compute, makes it an attractive competitor to the McKean-Hettmansperger test.

Educational researchers commonly examine their data for evidence of problems, nonnormality for example, that can threaten statistical conclusion validity. When scores are independently and identically (normally)-distributed with common variance σ^2, parametric tests are optimal for testing general linear model-based hypotheses. When normality does not hold, nonparametric tests represent an important data-analytic alternative to parametric tests (e.g., Lehmann, 1975, pp. 171-175; Marascuilo & McSweeney, 1977, p. 89; Zimmerman & Zumbo, 1993). (Other criteria for distinguishing between parametric and nonparametric tests have been formulated, see, e.g., Kendall & Stuart, 1979, pp. 497-498; Marascuilo & McSweeney, 1977, pp. 3-6). Comparisons of parametric and nonparametric estimators and tests under realistic data conditions, such as small sample sizes and nonnormal data, have spawned a considerable research literature. Despite its size, this literature is rather narrow in that its focus has been on relatively simple experimental designs and analyses (e.g., two groups).

Serlin and Harwell (2001) argued that nonparametric methods are under-used in educational research, in part because educational researchers are not aware of nonparametric tests that are available for complex experimental designs and analyses, such as factorial designs and multiple regression analyses. They also suggested that many educational researchers are not aware that such analyses can be performed using existing computer programs, such as SPSS (SPSS Inc., 1999) or Minitab (Minitab Inc., 2000). Serlin and Harwell concluded that the development of general linear model-based nonparametric procedures holds great promise for increasing the use of these nonparametric methods in educational research. They also pointed out that little is known about the behavior of these tests for realistic data conditions.

Serlin and Harwell indicated that three general linear model-based nonparametric procedures are especially promising. The aligned-rank procedure of Puri and Sen (1971, 1985) tests hypotheses about parameters of interest after eliminating so-called nuisance parameters. This involves aligning the raw scores using estimates of the nuisance parameters, ranking the aligned values (residuals), and computing a test statistic that follows a chi-square distribution. The rank-based procedure of McKean and Hettsmansper (1976) and Hettmansperger and McKean (1977) involves comparing two models (i) a reduced model containing nuisance parameters is fitted to the raw data and the residuals obtained and ranked (ii) the ranked and unranked residuals are used to compute a measure of dispersion (iii) steps (i) and (ii) are repeated for a full model containing the nuisance parameters and the parameters of interest (iv) a test statistic is computed based on the difference in the dispersion measures for the reduced and full models. Another promising procedure is the aligned-rank-transform described in Fawcett and Salter (1984), in which the rank-transform procedure of Conover and Iman (1976) is applied to data that have been aligned for nuisance parameters. Fawcett and Salter did not consider the aligned-rank-transform in a general linear model context, but doing so greatly extends the use of this method.

Theoretically, these procedures allow educational researchers to perform nonparametric tests for data obtained from complex experimental designs and data-analytic models using existing data analysis software. Estimators and tests associated with the Puri and Sen (1971, 1985) and McKean and Hettmansperger (1976) procedures have similar or identical properties asymptotically, and there is some evidence that the asymptotic properties of the aligned-rank-transform are similar to those of Puri and Sen and McKean and Hettmansperger. In any event, the literature on their performance for the less-than-asymptotic case is quite sparse. For example, available evidence that aligned-rank tests are excellent competitors to their parametric counterparts for controlling Type I errors at α and showing good statistical power (e.g., Adiche, 1978; Gorham, 1998; Puri & Sen, 1985) is limited to a small number of designs (e.g., randomized-block) and data analyses. Similarly, there is evidence that the McKean and Hettmansperger procedure performs reasonably well except when sample sizes are small, but available work has focused almost exclusively on factorial models and a few distributions. There has been even less study of the aligned-rank-transform procedure. In short, there is a substantial gap in the nonparametric literature of the behavior of these nonparametric tests for realistic data conditions. This paper reports the results of a computer simulation study of their behavior.

We first describe the three procedures and available theoretical and empirical evidence of their behavior. Next we introduce additional tests suggested by these procedures, and then report the results of a computer simulation study that investigated the behavior of the tests. We conclude by describing areas in which additional research is needed.

Puri and Sen's Aligned-Rank Test

Puri and Sen (1985, pp. 238-287) described a general linear model-based aligned-rank procedure, originally introduced by Mehra and Sarangi (1967) for main effects in additive models and extended by Sen (1968), that has several desirable properties. This procedure, which has its roots in the work of Hodges and Lehmann (1962), assumes that

$$F_i(y) = F(y_i - B_0 - \mathbf{B}'(\mathbf{x}_i - \bar{x})),\ i = 1,2,\ldots,N \tag{1}$$

underlies the data, where $F_i(y)$ is a continuous distribution function for the ith subject, y represents the dependent variable, B_0 is an intercept, $\mathbf{B}$ is a q x 1 vector of partial regression parameters consisting of q_1 nuisance parameters in the vector $\mathbf{B}_1$ and q_2 parameters of interest in $\mathbf{B}_2$ ($\mathbf{B} = \mathbf{B}_1, \mathbf{B}_2$, $q = q_1 + q_2$), $\mathbf{x}_i$ is a q x 1 vector of fixed and known predictor values for the ith subject, and $\bar{x}$ is a q x 1 vector of means for the predictor variables.

The Puri and Sen model applies equally to data fitting the correlation model

$$F_i(y|\mathbf{x}) = F(y_i - B_0 - \mathbf{B}'(\mathbf{x}_i - \bar{x})) \qquad (2)$$

in which the predictors are random and $F_i(y|\mathbf{x})$ represents the conditional distribution of y given $\mathbf{x}$ (Sampson, 1974). The assumption of continuity of the F_i theoretically eliminates the problem of tied scores, but when these occur conventional practice is to assign midranks. As long as the proportion of ties in the raw data is relatively small the midranks will have a negligible effect on the test (Lehmann, 1975, p. 18).

To test the hypothesis Ho: $\mathbf{B}_2 = \mathbf{0}$, a regression of y on the q_1 predictors associated with the nuisance parameters is performed and the resulting residuals are computed using $y_i - \mathbf{x}_i\mathbf{b}_1$, where $\mathbf{b}_1$ is a vector of estimated slopes. Either ordinary least squares (OLS) parameter estimates or rank estimates can be used to generate these residuals. Aubuchon and Hettmansperger (1984) pointed out that the impact of OLS versus rank estimates on tests in small samples is unknown, but most authors (e.g., Hettmansperger & McKean, 1998; Puri & Sen, 1985; Adiche, 1984) employ OLS estimates.

Under model assumptions the residuals are free of the effects of the q_1 nuisance parameters. They are then ranked (say, R_i). Next, the linear rank statistics $\hat{L}_k$ (Puri & Sen, 1985, p. 247) are computed for each predictor/dependent variable pairing for the q_2 predictors associated with the regression parameters of interest:

$$\hat{L}_k = \sum_{i=1}^{N} (x_{ik} - \bar{x}_k)[R_i], \; k = 1,2, .., q_2 \qquad (3)$$

Using the original x_{ik} values in equation (3) is an application of what Puri and Sen call the mixed-rank model; ranking the x_{ik} and using these ranks in the analysis is an application of the pure-rank model. These models lead to estimators and tests with the same properties, and we focus on the mixed-rank model.

It is clear in equation (3) that the $\hat{L}_k$ are proportional to the slopes, and that centering the predictors produces a nonparametric analogue of the least squares normal equations. This proportionality means that researchers do not need to compute the $\hat{L}_k$ and can instead simply compute slopes with the usual OLS expressions. The $\hat{L}_k$ (or slopes) are highly efficient compared to the usual least-squares estimators for a normal distribution and are asymptotically normal, clearing the way for an omnibus test statistic that follows a chi-square distribution (Puri & Sen, 1985, chpt. 7). Another advantage of the $\hat{L}_k$ is that they are robust compared to the usual least squares estimators that minimize $\sum (y_i - \mathbf{x}_i\mathbf{b})'(y_i - \mathbf{x}_i\mathbf{b})$, because the effect of outliers enters in a linear rather than a quadratic fashion (Draper, 1988).

Puri and Sen (1985, p. 247) proposed an aligned-rank test (PSAR) based on the $\hat{L}_k$ and their asymptotic variances that can be written in the form

$$\text{PSAR} = (N\text{-}1)\ \theta \sim \chi^2_{q2}(1\text{-}\alpha) \tag{4}$$

where θ represents a measure of explained variation. The PSAR test is asymptotically distribution-free and is asymptotically distributed as a chi-square variable with q_2 degrees of freedom. (The PSAR test is identical to the test proposed by Adichie (1978) for the single predictor case). In a regression model θ equals SSRegression/SSTotal, where SSRegression = $\mathbf{b'_2ssx_2b_2}$, $\mathbf{b_2}$ is a q_2 x1 vector of estimated slopes for the parameters of interest, $\mathbf{ssx_2}$ is a q_2 x q_2 sum of cross-products matrix for the predictors associated with the parameters of interest, and SSTotal = $\sum_{i=1}^{N} (R_i - \overline{R})^2$, where $\overline{R}$ is the overall mean of the ranks. As illustrated in Serlin and Harwell (2001), the PSAR test can easily be computed using existing software.

The assumptions underlying the PSAR test are that the y_i are continuous, independently and identically distributed, and that the sample size is large enough to ensure the validity of probabilistic inference based on a chi-square distribution. Under the assumption that the y_i follow a normal distribution, the A.R.E. of the PSAR test using ranks compared to the normal-theory likelihood ratio test is approximately .96, and the A.R.E. of the PSAR test for normally-distributed data with a normal-scores transformation compared to the likelihood ratio test is one (Puri & Sen, 1985, pp. 251-252).

Although contrasts for the B_k are available (Puri & Sen, 1985, chpt. 6), the PSAR procedure does not permit model-checking that is common in regression (e.g., studentized statistics). Aligned-rank procedures like PSAR have also been criticized on the grounds that their performance is suboptimal in certain settings. For example, Hettmansperger and McKean (1983) described computer simulation results for a test of parallel slopes in which an aligned-rank test showed inflated Type I error rates.

Research for the PSAR Test

Studies of the PSAR test have overwhelmingly focused on the factorial case. Akritas (1990) provided some A.R.E. comparisons between a rank-transform test and the PSAR test for a completely between-subjects factorial design. These results demonstrated that the A.R.E. of the usual rank-transform F-test was higher than the PSAR test for a normal distribution, but the A.R.E. of PSAR exceeded that of the rank-transform F-test for logistic and double exponential distributions. However, Brunner and Dette (1992) argued that the rank transform test considered by Akritas (1990) was not really a rank- transform test, but instead was a test in which the ranks are divided by an estimated standard deviation and then substituted into the usual F-test.

Harwell (1991) used a simulation study to examine the behavior of the PSAR test. For a 3 x 2 design with various cell sizes and distributions, Harwell found that as long as cell sample sizes were at least 8 the PSAR test controlled its Type I error rate and showed good power compared to the F-test and some nonparametric competitors; for smaller cell sample sizes the Type I error rates were frequently inflated. Toothaker and Newman (1994) also used a factorial design and found that the PSAR test had inflated Type I error rates for cell sample sizes of 5; for larger samples sizes the test controlled its error rate at the nominal level. Other simulation studies investigating the PSAR test include Conover and Iman (1976), Harwell and Serlin (1994), McSweeney (1967), Salter and Fawcett (1993), and Yohai and Ferretti (1987). The general result of these studies is that the Type I error rate differed noticeably from the nominal value for small sample sizes, for example, 3-5 cases per cell in a factorial design, with the test sometimes performing better for certain distributions. For larger sample sizes, the PSAR test generally did a good job of controlling its Type I error rate and showed good power.

McKean and Hettmansperger's Rank-Based Test

Building on the work of Jaeckel (1972), McKean and Hettmansperger (1976, 1977) described a two-step modeling procedure in which a sum of products of the ranked and unranked residuals for a reduced model containing nuisance parameters are compared to a sum of products for the ranked and unranked residuals for a full model containing the parameters of interest plus the nuisance parameters. The McKean-Hettmansperger method is similar to the usual least-squares model-fitting procedure, and supports model-checking procedures and contrasts for the B_k (McKean, Sheather, & Hettmansperger, 1990).

The test is based on a reduction in residual dispersion (RD) assuming that the model in equation (1) or (2) underlies the data. McKean and Hettmansperger defined the linear rank statistic

$$D_i(y_i - \mathbf{x}_i\mathbf{b}) = \sum_{i=1}^{N} (y_i - \mathbf{x}_i\mathbf{b}) [R_i^w(y_i - \mathbf{x}_i\mathbf{b})], \quad (5)$$

where R_i^w represents Wilcoxon scores generated from the score function $\varphi = (12)^{1/2}(R_i/(N+1) - 1/2)$. Use of the Wilcoxon scores ensures that $D_i(y_i - \mathbf{x}_i\mathbf{b})$ in equation (5) does not depend on the intercept (Draper, 1988). Hettmansperger and McKean (1998, p. 163) showed that the statistic in equation (5) is asymptotically unbiased, and if the population distribution is symmetric it is unbiased for any sample size. They also claimed that their statistic was resistant to the effects of outliers. However, Puri and Sen (1985, p. 282) noted that the statistic in equation (5) is more susceptible to outliers than the statistic in equation (2) because of the use of the residuals $(y_i - \mathbf{x}_i\mathbf{b})$.

Solving equation (5) for the slopes produces a rank analogue of the normal equations, which do not have a closed solution except in the case of $q = 1$. Thus, estimating the slopes requires an iterative technique and specialized software (Hettmansger & McKean, 1998, pp. 184-189). Draper (1988) pointed out that the estimated slopes are not necessarily unique because they may reflect one of several minima, but indicated that experience has suggested this is rarely a problem. The result is a fitted rank-based regression model with q slopes that are highly efficient compared to the usual least-squares estimators for a normal distribution.

The RREGRESS command in Minitab (Minitab Inc., 2000) will estimate the slopes that minimize the expression in equation (5), as will the interactive RANOVA program maintained by J. McKean at the web address http://www.stat.wmich.edu/slab/RGLM/index.htm. These slopes have an approximate distribution of $N(B_k, \tau^2 (\mathbf{x}'\mathbf{x})^{-1})$, permitting confidence intervals to be constructed about the B_k (Hettmansperger & McKean, 1998, p. 189). If a sample intercept is desired Aubuchon and Hettmansperger (1984) recommended using the median of the residuals, although if the distribution of residuals is assumed to be symmetric the median can be estimated using ranks along with the slopes (McKean

& Hettmansperger, 1978).

To adapt the statistic in equation (5) to test hypotheses McKean and Hettmansperger (1976) employed the following strategy: Suppose we wish to examine the contribution of $\mathbf{x}_2$ to explaining variation in y after taking into account the contribution of $\mathbf{x}_1$. The McKean-Hettmansperger procedure begins by fitting a reduced model of the form $\hat{y}_i = b_{01} + \mathbf{x}_1 \mathbf{b}_1$ to the y_i, where b_{01} is an estimated intercept and $\hat{y}_i$ represents fitted values. Then $D_i(y_i - \mathbf{x}_i\mathbf{b})$ in equation (5) is computed for the reduced model. Next a full model of the form $\hat{y}_i = b_{02} + \mathbf{x}_1\mathbf{b}_1 + \mathbf{x}_2\mathbf{b}_2$ is fitted to the data and the statistic in equation (5) is computed a second time.

McKean and Hettmansperger (1976) suggested that the rank-based statistic,

$$\text{MHCHI} = \text{RD}/(\hat{\tau}/2) \tag{6}$$

be used to test Ho: $\mathbf{B}_2 = 0$. In equation (6), $\text{RD} = D_i(y_i - \mathbf{x}_{1i}\mathbf{b}_1)_{\text{reduced model}} - D_i(y_i - \mathbf{x}_{1i}\mathbf{b}_1 - \mathbf{x}_{2i}\mathbf{b}_2))_{\text{full model}}$ and $\hat{\tau}$ is an estimate of a scale parameter τ, similar to the least squares parameter σ. The McKean-Hettmansperger chi-square test (MHCHI) is then compared to a chi-square value with q_2 degrees of freedom. Computer simulation studies by Hettmansperger and McKean (1977) and Draper (1981) suggested that the test statistic in equation (6) be modified to the form

$$\text{MHF} = (\text{RD}/q_2)/(\hat{\tau}/2) \tag{7}$$

and compared to a F critical value with q_2 and $N - q - 1$ degrees of freedom. The McKean-Hettmansperger F (MHF) and MHCHI tests are consistent, asymptotically distribution-free, and have the same A.R.E. under normality as the PSAR test (McKean & Hettmansperger, 1998, pp. 175-178).

A key feature of the tests in equations (6) and (7) is that they require estimates of τ. The parameter τ is used in several settings in nonparametric procedures, for example, in efficiency evaluations and to rescale tests so that they follow a known distribution (Hettmansperger, 1984, p. 244)). There are various ways to estimate τ (Revesz, 1984), but we focus on the two that are available in the RREGRESS command in Minitab (Minitab Inc., 2000). One is a Lehmann-type estimator based on the standardized length of a 90% Wilcoxon confidence interval (McKean & Hettmansperger, 1976). The other method used to estimate τ in RREGRESS is based on kernel estimation. (Details of how τ was estimated in our simulation study appear in Appendix A). Unfortunately, there is currently no widely available software that will perform the MHCHI or

MHF tests, but the RD statistic can be calculated with existing regression software in programs such as Minitab (Minitab Inc., 2000) and SPSS (SPSS Inc., 1999).

Research for the McKean-Hettmansperger Tests

Several computer simulation studies of the behavior of the MHF test have been done but apparently all have been quite limited in scope, almost always involving a focus on Type I error rates for a few conditions in a two-factor design. McKean and Sievers (1989) reported that the MHF test maintained its Type I error rate near the nominal value (with one exception) in an unbalanced 3x3 design with interaction for two heavy-tailed distributions (logistic, log-Pareto). Hettmansperger and McKean (1977) examined the MHF test for a balanced 3x3 design, small cell sample sizes (3, 5), and a double-exponential distribution, and reported that the estimated Type I error rates were consistently inflated for small samples (5-10). Hettmansperger and McKean (1983) performed a computer simulation study that used MHF to test for parallelism of slopes for three groups, sample sizes of 5 or 10, and double-exponential and Cauchy distributions. The MHF test maintained its Type I error rate near the nominal value for all conditions examined, and showed good power compared to some nonparametric competitors. Other studies of the MHF test include McKean and Hettmansperger (1978), Sievers and McKean (1986), McKean and Sheather (1991), McKean, Vidmar, and Sievers, (1989), Sievers and McKean, (1986), and Hettmansperger and McKean (1977). The general finding from these studies was that the MHF test maintained its Type I error rate unless cell sample sizes were quite small, and that the test frequently performed well for heavy-tailed distributions.

Aligned-Rank-Transform Test

The rank-transform procedure introduced by Conover and Iman (1976) requires that scores be ranked and submitted to a parametric test. This procedure is known to work well for tests in simple designs but less well for many complex designs, for example, factorial designs (Akritas, 1990). Salter and Fawcett (1984) provided a possible solution to this problem for a randomized block design by suggesting that the rank-transform be applied to data that have been aligned for nuisance parameters, producing an aligned-rank-transform test. Akritas (1991) pushed this notion further by suggesting that an aligned-rank-transform test could be applied to subhypotheses of the general linear model. The aligned-rank-transform test is the same as the PSAR test except that the test statistic is divided by q_2 to produce an (approximate) F that is then compared to a critical F value with q_2 and $N - q - 1$ degrees of freedom. The test is easily computed with available data analysis software.

Research for the Aligned-Rank-Transform Test

Mansouri and Chang (1995) provided theoretical results for the ART test and also reported simulation results for a factorial design that showed the test controlled its Type I error rate at α . Mansouri (1998) provided the limiting distribution and A.R.E. of an ART test for a balanced incomplete blocks design. Kepner and Wackerly (1996) provided A.R.E. comparisons for an ART test for a balanced incomplete repeated measures design using Wilcoxon rank scores, and showed that the test was particularily attractive for heavy-tailed distributions. Akritas (1993) presented an aligned-rank-transform test that can be used when data are heteroscedastic.

Salter and Fawcett (1984) performed a computer simulation study that provided evidence that the F distribution could provide a satisfactory approximation to the distribution of their aligned-rank-transform (ART) test in a randomized block design. Salter and Fawcett (1993) studied the ART test for a completely between-subjects factorial design with varying sample sizes and distributions, and found that the test controlled its Type I error rate and showed good power for cell sample sizes greater than 10. Other studies of the ART test have been reported by Groggel (1987), Harwell and Serlin (1994), and Gorham (1998). The ART test frequently produced conservative Type I error rates for small samples, with corresponding low power; for larger sample sizes, however, the test typically performed well.

These results suggest that an ART test may be an important competitor to the PSAR and MHCHI/MHF tests. However, there do not appear to be any studies available of an ART test in hierarchical regression.

Before continuing, it is important to point out that the statistical hypotheses tested by the various nonparametric tests may differ from one data analysis to another. Recall that the PSAR and MHCHI/MHF tests test the hypothesis Ho: $\mathbf{B}_2 = \mathbf{0}$. Akritas and Arnold (1994) pointed out that rejection of this Ho does not necessarily imply that the slopes do not equal zero unless additional assumptions are imposed on the data to ensure that rejection is attributable to nonzero slopes and not to other distributional characteristics such as scale, skewness, and/or kurtosis.

Akritas and Arnold argued that nonparametric hypotheses should be defined in a way that does not place additional assumptions on the data, such as by writing Ho in terms of distribution functions. For example, they would replace

Ho: $\mathbf{B}_2 = \mathbf{0}$ with Ho: $F_i(y) = F(y_i|(\mathbf{x}_i - \bar{x})$, with the latter Ho described as fully nonparametric because it does not directly depend on any parameters (Akritas & Arnold, 1994). Under the null hypothesis, Ho: $\mathbf{B}_2 = \mathbf{0}$ and Ho: $F_i(y) = F(y_i|(\mathbf{x}_i - \bar{x})$ are identical but under H_1 they differ because the fully nonparametric version does not directly attribute the rejection to nonzero slope parameters. Following the lead of Puri and Sen (1971, 1985), Adiche (1978), Hettmansperger (1984), and others we assume that all of the nonparametric tests test Ho: $\mathbf{B}_2 = \mathbf{0}$. Clearly, the data must be examined for evidence supporting the additional assumptions associated with writing Ho in this fashion (Hettmansperger & McKean, 1998, p. 234).

Serlin-Harwell Aligned-Rank Procedure (SHARP)

An examination of the strategies behind the PSAR, MHCHI, MHF, and ART tests suggests other nonparametric tests that can be constructed to examine the effects of a set of variables after the contribution of other variables has been removed. Recall that in the PSAR test the first step is to create residuals, which are free of the effects of the nuisance variables, and in the McKean-Hettmansperger procedure a function of the ranks of residuals from the full and reduced models are compared. We consider a marriage of these strategies.

Suppose we wish to examine the effects of the predictors x_3-x_4 on y after the effects of x_1-x_2 were taken into account. Suppose also that x_1-x_2 were used to predict y, the residuals obtained and ranked, and a sum of squares regression ($SSReg_{\text{reduced model}}$) obtained by using x_1-x_2 to predict the ranked residuals. Computing $SSReg_{\text{reduced model}}/SSTotal$ then produces $R^2_{\text{reduced model}}$. Next we predict the ranked residuals using x_1-x_4 and compute $R^2_{\text{full model}}$. The hypothesis Ho: $\mathbf{B}_2 = 0$ can be tested using

$$SHARPCHI = (N-q_1-1)[(R^2_{\text{full model}} - R^2_{\text{reduced model}})/(1 - R^2_{\text{reduced model}})] \sim \chi^2_{q2} \quad (8)$$

The $(N-q_1-1)$ are the degrees of freedom associated with the sum of squares left over after the ranked residuals are predicted from the reduced model. Dividing SHARPCHI by q_2 produces a statistic with an F distribution with q_2 and N-q-1 degrees of freedom (SHARPF). We note that Mansouri (1996) proposed a similar test of the form $q_2[(SSError_{\text{reduced model}} - SSError_{\text{full model}})/MSError_{\text{full model}}]$, which follows a chi-square distribution with q_2 degrees of freedom.

The various tests have many similarities. The PSAR, MHCHI, and SHARPCHI tests are based on a quadratic form in the ranks that are asymptotically distributed as a chi-square variable, and can easily be converted to F-tests. In addition, the tests produce the same or similar results when applied to data from simpler experimental designs (e.g., a single-factor design); however, for more complex designs there may be sharp differences in their behavior. There may also be differences in their behavior for small sample sizes and particular distributional forms.

Simulation Study

As indicated previously, literature on the performance of general linear model-based nonparametric tests has largely been limited to analyses based on factorial designs. The Type I error and power performance of the PSAR, MHCHI, MHF, ART, SHARPCHI, and SHARPF tests were examined for realistic data conditions for the hierarchial regression model. We chose hierarchial regression to compare the nonparametric tests for two reasons. First, this procedure has been regularly used in educational research, and, second, this study will add to the nonparametrics literature because there is apparently no evidence of the behavior of these tests for the hierarchial regression model.

We used a computer simulation study to generate data for a hierarchial regression model, which in turn were used to examine the Type I error and power behavior of the various tests. Although an analytic approach to the behavior of these tests is preferred because of its generalizability, available theoretical work for the nonparametric general linear model-based tests investigated in this study, when such results exist, assumes that quite large sample sizes are present. Since large sample sizes may not occur in practice, it is important to study a test's behavior under realistic conditions, such as small sample sizes and different distributions (Draper, 1988). We used traditional rank and Wilcoxon scores for the various nonparametric tests, although we acknowledge that choice of scoring functions is important because different score functions give rise to estimators and tests with different properties (Draper, 1988; Naranjo & McKean, 1997; Policello & Hettmansperger, 1976).

The hierarchial regression model in the simulation had a total of four predictors: x_1-x_2 represented the reduced model and x_1-x_4 the full model. The statistical null hypothesis tested was Ho: $\mathbf{B}_2 = 0$. The design factors of the simulation study were (a) Distribution of the residuals (normal with skewness (γ_1) and kurtosis (γ_2) of 0; chi-square with 8 df ($\gamma_1 = 1$, $\gamma_2 = 1.5$); chi-square with 4 df ($\gamma_1 = 1.41$, $\gamma_2 = 3$); approximate Cauchy ($\gamma_1 = 0$, $\gamma_2 = 25$)), (b) Sample size (N = 20, 40, 60, or 80 representing sample size to number of predictor ratios of 5:1, 10:1, 15:1, and 20:1, respectively, for Type I error rate runs; N=20, 60 for power runs); (c) $\rho_{x1x2} = \rho_{x3x4} = 0.0$ or 0.3. In all cases the residuals were homoscedastic and the ρ_{x1x3}.

$= \rho_{x1x4} = \rho_{x2x3} = \rho_{x2x4} = 0.3$.

The selection of sample sizes and distributions was made based on conditions observed in the educational research literature. The chi-square distributions with 8 and 4 degrees of freedom represent increasingly skewed and kurtic data, whereas the approximate Cauchy represents an extremely heavily-tailed distribution and is important to examine because of theoretical and empirical evidence that rank-based tests are often superior to parametric tests for such distributions. The $\rho_{x1x2} = \rho_{x3x4}$ correlation of 0 represents the ideal case of no overlapping variation among predictors, whereas $\rho_{x1x3} = \rho_{x1x4} = \rho_{x2x3} = \rho_{x2x4} = .30$ means there was some shared variation among predictors but not enough to raise concerns about collinearity. (The same logic guided our selection of $\rho_{x1x3} = \rho_{x1x4} = \rho_{x2x3} = \rho_{x2x4} = .30$). Assuming a normal distribution, the data were generated such that the x_1-x_2 predictors accounted for 20% of the variance in y; the remaining predictors (x_3-x_4) were added to the model and accounted for differing amounts of additional variance in y (expressed through correlations $\rho_{yx3} = \rho_{yx4}$) needed to achieve a theoretical power of .70 for varying sample sizes. When the partial correlation $R_{yx3x4.x1x2}$ equaled 0, rejections of Ho: $\mathbf{B}_2 = 0$ counted toward the estimated Type I error rate; for the non-zero case rejections counted towards the estimated power.

<u>Data Generation</u>

The following steps were taken to generate data with the desired characteristics (1) 5*N scores following a multivariate-normal distribution were generated using the Kaiser and Dickman (1962) procedure. These values were then transformed to the various nonnormal distributions following the Vale and Maurelli (1983) procedure, which combines the Kaiser and Dickman (1962) and Fleishman (1978) procedures. Evidence of the success of the data generation for various distributions is provided in Table 1 for two representative same sizes (N = 20, 60). Details on the computation of correlations among the x_k and y variables used to generate estimated Type I error rates and power values is given in Appendix B.

Overall, the design of the simulation involved 4 (distribution) x 4 (sample sizes for Type I error rate runs) x 2 (correlations with pairs of predictors) + 4 (distribution) x 2 (sample sizes for power runs) x 2 (correlations with pairs of predictors) = 48 conditions. Twenty-thousand replications per condition were used to estimate the Type I error rate and power of the tests, ensuring that estimates showed little sampling error. For each simulated dataset, the PSAR, MHCHI, MHF, ART, SHARPCHI, SHARPF, and parametric F-tests were calculated and compared to the appropriate critical value for the null case, producing a proportion of rejections ($\hat{\alpha}$) for each condition studied (In all cases the nominal Type I error rate was .05). Both the confidence interval and window methods for estimating τ described in Appendix B were used. As a result, each McKean-Hettmansperger test was computed using a confidence interval-based estimate as well as a window-based estimate. A similar process was followed for the power case (1- $\hat{\beta}$). The simulation program was written in Microsoft Fortran 4.0.

Results

Type I Error Results

Summaries of the Type I error performance of the tests are displayed in Table 2. The estimated Type I error rates of the McKean-Hettmansperger tests using window estimates of τ were very similar to those produced by the tests based on confidence interval estimates, and only the latter are presented. Table 2 shows that the MHF test using a confidence interval estimate for τ (MHFCI) produced an average estimated Type I error rate closest to the nominal value of .05 (.0517). Next closest is the SHARPCHI test with an estimated error rate of .0535, followed, in order, by PSAR (.0449), SHARPF (.0392), MHCHI chi-square test using a confidence interval estimate for τ (MHCHICI, .0662), ART (.0334), and the parametric F-test (.0675). The MHFCI, SHARPCHI, and the PSAR tests overall provided satisfactory control of Type I error rates, with the remaining tests providing somewhat less control. As expected, for a normal distribution the average Type I error rate of the parametric F-test was closest to the nominal value (.0502), followed by the MHFCI (.0468), SHARPCHI (.0450), MHCHICI (.0612), SHARPF (.0329), PSAR (.0304), and ART (.0212) tests.

Table 3 reports the estimated Type I error rates for each test by distribution and sample size (Evidence described below indicated that the correlation within pairs of predictors of 0.0 or .30 did not have much effect on error rates and this variable was not included in Table 3). Many of the estimated error rates were reasonably close to .05 and indicated that several tests often showed adequate control of error rates. However, for purposes of recommending one or more of these tests, we further characterized their control of Type I error rates through the use of four categories: Estimated $\hat{\alpha}$ values within the range .05 ± 2SD [SD = $\{\alpha(1-\alpha)/20{,}000)]^{1/2}\}$ = .0470 to .0530 were considered to represent excellent control; values between .05 ± 2-3 SD (.0453 to .0469, .0531 to .0546) were characterized as mildly inflated or conservative error rates that represented good error rate control (identified in Table 3 with an *); values between .05 ± 3-4 SD (.0438 to .0452, .0547 to .0561) were characterized as mildly to moderately inflated or conservative error rates that represented adequate control and are indicated by a +; values outside .05 ± 4SDs (< .0438 or > .0561) represented more pronounced inflation or overly conservative values that reflected unsatisfactory control of error rates and are indicated by a &. These categories allowed us to further discriminate tests showing excellent control of Type I error rates from those showing less satisfactory control (Recall that the use of 20,000 replications means that each estimated error rate in Table 3 should be close to that test's "true" α for the conditions studied).

An examination of Table 3 shows that for a normal distribution the F-test, as expected, produced $\hat{\alpha}$ values close to .05 regardless of sample size, providing further evidence of the credibility of the simulation. Across the full set of conditions reported in Table 3, the PSAR, ART, MHCHICI, F, and SHARPF tests showed unsatisfactory control of error rates for more than half of the conditions studied, with the ART and PSAR tests consistently producing overly conservative values and the others inflated values. Unlike some previous simulation studies, the McKean-Hettmansperger tests were sensitive to the Cauchy distribution. The MHFCI test, on the other hand, did a good job of controlling Type I error rates, followed by the SHARPCHI test which showed adequate control except for the Cauchy distribution. Still, the number of independent variables in the simulation design means that complex effects in the estimated error rates, such as whether the relationship between error rates and distribution depends on sample size, may be present but are not immediately discernible in these data. We next attempted to tease out this information.

Following the advice of Hoaglin and Andrews (1975) to analyze data from simulation studies for evidence of important patterns, we fitted three-way ANOVA models to the estimated Type I error rates for each test to determine which effects appeared to be the largest and whether interactions needed to be considered in the interpretation. We first transformed the $\hat{\alpha}$ using an arcsin transformation described in Marascuilo and McSweeney (1977, pp. 147-148) that produces values whose mean and variance are independent of one another, and whose sampling distribution is quickly approximated by a normal distribution. Because the design was unreplicated we did not model the three-way interaction, which allowed within-cell variance estimates to be obtained. Model-checking revealed no strong departures from normality or homogeneity of variance. Each effect was tested using a nominal Type error rate of .05.

The ANOVA results are summarized in Table 4 in the form of $\hat{\eta}^2$ statistics, defined as the sum of squares of a statistically significant effect over the sum of squares total. We also computed $\hat{\omega}^2$ statistics (Hays, 1973), which are less biased than $\hat{\eta}^2$ statistics as measures of effect size. The $\hat{\omega}^2$ statistics were similar to the $\hat{\eta}^2$ values, however, and only the latter are reported in Table 4. The majority of the two-way interactions were not statistically significant, and among the handful that were the associated $\hat{\eta}^2$ never exceeded .048. As a result, our focus was on the main effects, and Table 4 reports $\hat{\eta}^2$ values for the Distribution, Correlation Within Pairs of Predictors (x_1-x_2, x_3-x_4), and Sample Size effects. The small $\hat{\eta}^2_{\text{Within-Pair}}$ values provide a rationale for not including this factor in reporting the estimated Type I error rates in Table 3.

Table 4 indicates that the MHFCI (.83) test was quite sensitive to the underlying distribution, along with the PSAR test (.47). The SHARPF (.08) test, on the other hand, showed noticeably less sensitivity to distribution. The estimated error rates of several tests (SHARPF, MHCHICI, ART, SHARPCHI) were also sensitive to sample size, which was expected because several of the tests have been shown to have an asymptotic error rate of α. The fact that the PSAR (.20) test was less sensitive to sample size was unexpected.

The sensitivity of the tests to sample size was explored further by re-running the ANOVAs with Sample Size restricted to the N = 60 or 80 cases, which should reflect the asymptotic behavior of $\hat{\alpha}$ more than N = 20 or 40. That is, re-running the ANOVAs with N = 60 or 80 should shrink the $\hat{\eta}^2_{\text{Sample Size}}$ values of the tests. The values in parentheses in Table 4 for $\hat{\eta}^2_{\text{Sample Size}}$ are for the N = 60 or 80 case and provide evidence that the tests behaved as expected theoretically, with all nonparametric tests showing substantial to huge decreases in $\hat{\eta}^2$ when using larger sample sizes.

In sum, the results reported in Tables 2-4 suggest that the parametric F-test be used for normally-distributed data, as predicted by theory. For the nonnormal distributions studied the performance of the MHFCI test was superior to the others, followed by the SHARPCHI test. The PSAR and ART tests, on the other hand, performed poorly for most conditions studied. The remaining tests showed a mixed Type I error rate pattern, doing well for some conditions and less well for others.

Power Results

Summaries of the overall power performance of the nine tests are displayed in Table 2. For a normal distribution the average power values were MHCHICI (.7094), F (.6985), MHFCI (.6588), SHARPCHI (.6617), and SHARPF (.5653). Because of the difficulty of interpreting power values when Type I error rates stray from the nominal value (.05), we focus on conditions in which the tests did not show pronounced inflation ($\hat{\alpha} \leq .0562$).

Table 5 reports individual power values (Power rates associated with $\hat{\alpha} > .0562$ are indicated with a &). For a normal distribution, the estimated power values of .6984 and .6909 for the F-test for N = 20 and 60, respectively, provided further evidence of the credibility of the simulation. For the moderately and strongly skewed/kurtic chi-square distributions the MHFCI produced the highest power, with values fairly close to .70, followed by the SHARPCHI and SHARPF tests. The ART and PSAR tests performed poorly, which is not surprising given their conservative Type I error rates for most conditions. Factorial ANOVAs were not run for the power values because removing those values associated with an inflated $\hat{\alpha}$ leaves a substantially unbalanced design.

Summary

The simulation results suggest the following conclusions (1) As predicted by theory, the parametric F-test should be used for normally-distributed data regardless of sample size. (2) For the nonnormal distributions studied, the McKean-Hettmansperger F-test using a confidence interval estimate of τ produced Type I error rates close to .05 and showed good power even for smaller sample sizes, followed by the Serlin-Harwell aligned-rank chi-square test. (3) The performance of the Serlin-Harwell aligned-rank F-test was mixed, while those of the Puri and Sen, aligned-rank-transform, and McKean-Hettmansperger chi-square test with a confidence interval estimate were generally poor.

On balance, the top-performing test for the nonnormal distributions studied was the McKean-Hettmansperger F-test using a confidence interval estimate. The Serlin-Harwell aligned-rank chi-square test performed almost as well for many of the conditions studied, which, combined with the fact that it is easier to compute, makes it an attractive competitor to the McKean-Hettmansperger test.

Future Research

This study provides evidence of the behavior of several general linear model-based nonparametric tests in a hierarchial regression analysis under realistic conditions. Future work might include studying the behavior of the tests for other statistical procedures and conditions, such as factorial designs with heteroscedastic and nonnormal data. The result of this work will be the development of a literature that will provide educational researchers with credible nonparametric alternatives when analyzing data from complex research designs and data analyses.

References

Adichie, J.N. (1978). Rank tests of sub-hypotheses in the general linear regression. *Annals of Statistics*, *2*, 396-904.

Adiche, J.N. (1984). Rank tests in linear models. *In the Handbook of Statistics Vol 4* (P.R. Krishnaiah & P.K. Sen, Eds.). Amsterdam: North-Holland.

Akritas, M. G. (1990). The rank transform method in some two-factor designs. *Journal of the American Statistical Association*, *85*, 73-78.

Akritas, M.G. (1991). Limitations of the rank transform procedure: A study of repeated measures designs, Part I. *Journal of the American Statistical Association*, *86*, 457-460.

Akritas, M. G. (1993). Aligned rank tests for the linear model with heteroscedastic errors. *Journal of Statistical Planning and Inference*, *37*, 23-41.

Akritas, M.G., & Arnold, S.F. (1994). Fully nonparametric hypotheses for factorial designs I: Multivariate repeated measures designs. *Journal of the American Statistical Association*, *89*, 336-343.

Aubuchon, J.C., & Hettmansperger, T.P. (1984). On the use of rank tests and estimates in the linear model. In the *Handbook of Statistics* Vol 4, P.R. Krishnaiah and P.K. Sen (Eds.). New York: Elsevier.

Bean, S.J., & Tsokos, C.P. (1980). Developments in nonparametric density estimation. *International Statistical Review*, *48*, 267-287.

Brunner, E., & Dette, H. (1992). Rank procedures for the two-factor mixed model. *Journal of the American Statistical Association*, *87*, 884-888.

Chang, G., & Mansouri, H. (1993). Nonparametric analysis of covariance in randomized block designs. Unpublished manuscript.

Conover, W.J., & Iman, R.L. (1976). On some alternative procedures using ranks for the analysis of variance of experimental designs. *Communications in Statistics-Theories and Methods*, *14*, 1349-1368.

Draper, D. (1981). *Rank-based robust analysis of linear models*. Unpublished doctoral dissertation, University of California-Berkeley.

Draper, D. (1988). Rank-based robust analysis of linear models I: Exposition and review. *Statistical Science*, *3*, 239-271.

Fawcett, R.F., & Salter, K. (1984). A Monte Carlo study of the F test and three tests based on ranks of treatment effects in randomized block designs. *Communications in Statistics—Simulation and Computation*, *13*, 213-225.

Fleishman, A. (1978). A method for simulating nonnormal distributions. *Psychometrika*, *43*, 521-532.

Gorham, J.L. (1998). *The effects on Type I error rate and power of selected competitors to the ANOVA F test in the randomized block designs under non-normality and variance heterogeneity.* Unpublished dissertation, Rutgers, The State University of New Jersey.

Groggel, D.J. (1987). *Communication in Statistics-Simulation and Computation, 16*, 601-620.

Harwell, M.R. (1991). Completely randomized factorial analysis of variance using ranks. *British Journal of Mathematical and Statistical Psychology, 44*, 383-401.

Harwell, M.R., & Serlin, R.C. (1994). A Monte Carlo study of the Friedman test and some competitors in the single-factor, repeated measures design with unequal covariances. *Computational Statistics & Data Analysis, 17*, 35-49.

Hays, W.L. (1973). *Statistics for the social sciences* (2nd Ed.). New York: Holy, Rinehart, & Winston.

Hettmansperger, T.P. (1984). *Statistical inference based on ranks.* New York: Wiley.

Hettmansperger, T.P., & McKean, J.W. (1977). A robust alternative based on ranks to least squares in analyzing general linear models. *Technometrics, 19*, 275-284.

Hettmansperger, T.P., & McKean, J.W. (1983). A geometric interpretation of inferences based on ranks in the linear model. *Journal of the American Statistical Association, 78*, 885-893.

Hettmansperger, T.P., & McKean, J.W. (1998). *Robust nonparametric statistical methods.* London: Arnold.

Hoaglin, D.C., & Andrews, D.F. (1975). The reporting of computation-based results in statistics. *The American Statistician, 29*, 122-126.

Hodges, J.L., & Lehmann, E.L. (1962). Rank methods for combination of independent experiments in analysis of variance. *Annals of Mathematical Statistics, 33*, 482-497.

Jaeckel, L.A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics, 43*, 1449-1458.

Kaiser, H.F., & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika, 27*, 179-182.

Kendall, M. & Stuart, A. (1979). *The advanced theory of statistics Vol. 2* (4th Ed.). New York: Macmillan Publishing Company.

Kepner, J.L., & Wackerly, D. (1996). On rank transformations for balanced incomplete repeated-measures designs. *Journal of the American Statistical Association, 91*, 1619-1625.

Lehmann, E.L. (1975). *Nonparametrics: Statistical methods based on ranks.* San Francisco: Holden-Day.

Mansouri, M. (1996, August). Analysis of variance based on ranks using computing packages. Paper presented at the Joint Statistical Meetings, Chicago.

Mansouri, H., & Chang, G.H. (1995). A comparative study of some rank tests for interaction. *Computational Statistics and Data Analysis*, *19*, 85-96.

Marascuilo, L.A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole.

Maritz, J.S. (1995). Distribution-free statistical methods. London: Chapman and Hall.

McKean, J.W., & Hettmansperger, T.P. (1976). Tests of hypotheses based on ranks in the general linear model. *Communications in Statistics-Theory and Methods*, *8*, 693-709.

McKean, J.W., & Hettmansperger, T.P. (1978). A robust analysis of the general linear model based on one step R-estimates. *Biometrika*, *65*, 571-579.

McKean, J.W., & Sheather, S.J. (1991). Small sample properties of robust of linear models based on R-estimates: A survey. In *Directions in robust statistics and diagnostics, Part II*, W.Stahel and S. Weisberg (eds.). New York: Springer-Verlag.

McKean, J.W., Sheather, S.J., & Hettmansperger, T.P. (1990). Regression diagnostics for rank-based methods. *Journal of the American Statistical Association*, *85*, 1018-1028.

McKean, J.W., & Sievers, G.L. (1989). Rank scores suitable for the analysis of linear models under asymmetric error distributions. *Technometrics*, *31*, 207-218.

McKean, J.W., Vidmar, T.J., & Sievers, G.L. (1989). A robust two-stage multiple comparison procedure with application to a random drug screen. *Biometrics*, *45*, 1281-1297.

McSweeney, M. (1967). An empirical comparison of two proposed nonparametric tests for main effects and interaction. Doctoral dissertation, University of California-Berkeley.

Mehra, K.L., & Sarangi, J. (1967). Asymptotic efficiency of certain rank tests for comparative experiments. *Annals of Mathematical Statistics*, *38*, 90-107.

Minitab Inc. (2000). *Minitab for windows*. Minitab Inc.

Naranjo, J.D., & McKean, J.W. (1997). Rank regression with estimated scores. *Statistics & Probability Letters*, *33*, 209-216.

Policello, G.E., & Hettmansperger, T.P. (1976). Adaptive robust procedures for the one-sample location model. *Journal of the American Statistical Association*, *71*, 624-633.

Puri, M.L., & Sen, P.K. (1971). *Nonparametric methods in multivariate analysis*. Wiley: New York.

Puri, M.L., & Sen, P.K. (1985). *Nonparametric methods in general linear models*. Wiley: New York.

Revesz, P. (1984). Density estimation. In the *Handbook of Statistics* Vol 4, P.R. Krishnaiah and P.K. Sen (Eds.). New York: Elsevier.

Salter, K.C., & Fawcett, R.F. (1993). A robust and powerful rank test of interaction in factorial models. *Communications in Statistics-Simulation and Computation*, *22*, 137-153.

Sampson, A.R. (1974). A tale of two regressions. *Journal of the American Statistical Association*, *74*, 682-689.

Sen, P.K. (1968). On a class of aligned rank tests in two-way layouts. *Annals of Mathematical Statistics*, *39*, 1115-1124.

Serlin, R.C., & Harwell, M.R. (2001, April). A review of nonparametric tests for complex experimental designs in educational research. Paper presented at the annual meeting of the American Educational Research Association, Seattle.

Sievers, G.L., & McKean, J.W. (1986). On the robust rank analysis of linear models with non-symmetric error distributions. *Journal of Statistical Inference and Planning*, *13*, 215-230.

SPSS Inc. (1999). *SPSS for windows*. SPSS Inc.

Tapia, R.A., & Thompson, J.R. (1978). *Nonparametric probability density estimation*. Baltimore, MD: Johns Hopkins University Press.

Vale, C.D., & Maurelli, V.A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465-471.

Yohai, V.J., & Ferretti, N.E. (1987). Tests based on weighted rankings in complete blocks: Exact distribution and Monte Carlo simulation. *Communications in Statistics—Simulation and Computation*, *16*, 333-347.

Zimmerman, D. W., & Zumbo, B. D. (1993). Relative power of parametric and nonparametric statistical methods. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences, Volume 1: Methodological issues* (pp. 481-517). Hillsdale, NJ: Lawrence Erlbaum

Table 1

Evidence of the Success of the Data Generation

Sample Size	Distribution	Mean	Variance	Skewness	Kurtosis	($\rho_{x1x2} = 0$) $\bar{r}_{x1x2}$	($\rho_{x1x3} = .3$) $\bar{r}_{x1x3}$
	Normal (γ_1=0, γ_2=0)	-.0005	1.0006	.0007	.0018	.0001	.3011
N = 20	χ^2 (df=8) (γ_1=1, γ_2=1.5)	-.0001	1.0007	1.002	1.515	-.0022	.3000
	χ^2 (df=4) (γ_1=1.41, γ_2=3)	-.0001	.9988	1.412	2.984	-.0062	.2981
	Cauchy (γ_1=0, γ_2=25)	-.0006	.9985	-.0053	25.144	-.0017	.2998
	Normal	-.0004	.9999	.0008	-.0019	.0003	.3001
N=60	χ^2 (df=8)	-.0001	1.001	1.000	1.503	.0011	.3005
	χ^2 (df=4)	-.0003	.9990	1.412	2.985	.0001	.2991
	Cauchy	-.0017	1.001	-.0168	24.794	.0002	.3005

Each tabled value represents an average. The mean, variance, skewness, and kurtosis are based on data for N cases for 5 variables (4 predictors and y) across 20,000 replications for both Type I error rate and power conditions. For example, the first entry in the table of -.00054 is based on 20*5*20,000*2 = 4,000,000 scores. The results for the average correlation $\bar{r}_{x3x4}$ were similar to those for $\bar{r}_{x1x2}$, and the results for $\bar{r}_{x1x4}$ and $\bar{r}_{x2x4}$ were similar to those for $\bar{r}_{x1x3}$.

Table 2

Summary of Estimated Type I Error and Power Values for the Tests

Estimated Type I Error Values

	F test	Puri and Sen test	McKean-Hett mansperger chi square using interval estimate	McKean-Hett mansperger F using Interval estimate	SHARP chi square test	SHARP F test	Aligned rank transform test
Mean	.0675	.0449	.0662	.0517	.0535	.0392	.0334
Median	.0627	.0352	.0611	.0488	.0529	.0425	.0286
Maximum	.1015	.1520	.0967	.0703	.0804	.0718	.1365
Minimum	.0473	.0263	.0499	.0430	.0362	.0139	.0086
Std Deviation	.0156	.0274	.0128	.0077	.0099	.0156	.0273

Estimated Power Values

	F test	Puri and Sen test	McKean-Hett mansperger chi square using interval estimate	McKean-Hett mansperger F using Interval estimate	SHARP chi square test	SHARP F test	Aligned rank transform test
Mean	.6985	.3855	.7094	.6588	.6617	.5653	.2958
Median	.7009	.3987	.7192	.6587	.6505	.6044	.2905
Maximum	.7589	.5167	.7863	.7246	.7504	.6844	.4815
Minimum	.6617	.2071	.6337	.6155	.6063	.4355	.0994
Std Deviation	.0236	.0937	.0461	.0269	.0407	.0948	.1186

Table 3

Estimated Type I Error Values by Distribution and Sample Size

N	Distribution	F	PSAR	MHCHICI	MHFCI	ART	SHARPCHI	SHARPF
20	Normal	.0507	.0275&	.0757&	.0448+	.0104&	.0384&	.0148&
20	Moderately skewed/kurtic	.0589&	.0286&	.0794&	.0446+	.0102&	.0422&	.0148&
20	Skewed/kurtic	.0655&	.0309&	.0816&	.0498	.0106&	.0464*	.0170&
20	Cauchy	.0847&	.0275&	.0936&	.0632&	.0088&	.0473	.0178&
40	Normal	.0503	.0302&	.0594&	.0471	.0212&	.0456*	.0339&
40	Moderately skewed/kurtic	.0600&	.0342&	.0605&	.0474	.0249&	.0511	.0378&
40	Skewed/Kurtic	.0703&	.0400&	.0638&	.0513	.0290&	.0573&	.0431&
40	Cauchy	.0909&	.0615&	.0783&	.0656&	.0443+	.0631&	.0473
60	Normal	.0504	.0328&	.0565&	.0483	.0265&	.0482	.0405&
60	Moderately skewed/kurtic	.0613&	.0371&	.0572&	.0497	.0300&	.0550+	.0455*
60	Skewed/kurtic	.0693&	.0421&	.0562&	.0486	.0343&	.0599&	.0507
60	Cauchy	.0920&	.0925&	.0731&	.0641&	.0787&	.0697&	.0587&
80	Normal	.0495	.0310&	.0534*	.0470+	.0266&	.0477	.0422&
80	Moderately skewed/kurtic	.0614&	.0373&	.0513	.0464*	.0328&	.0537*	.0477
80	Skewed/Kurtic	.0700&	.0436&	.0525	.0469*	.0374&	.0597&	.0528
80	Cauchy	.0955&	.1212&	.0670&	.0614&	.1086&	.0715&	.0633&

* = outside 2SD (.0469, .0531), + = outside 3SD (.0453, .0546), & = outside 4SD (.0438, .0561). F = parametric F-test, PSAR = Puri and Sen aligned-rank test, MHCHICI = McKean-Hettmansperger chi-square test with confidence interval estimate of τ , MHFCI = McKean-Hettmansperger F test with confidence interval estimate, ART=Aligned-rank transform, SHARPCHI=Serlin-Harwell modified aligned-rank chi-square test, SHARPF = Serlin-Harwell modified aligned-rank F-test. Each tabled value is the proportion of rejections across 20,000 replications.

Table 4

Effect Size Estimates for ANOVAs of Estimated Type I Error Rates

Test	$\hat{\eta}^2_{\text{Distribution}}$	$\hat{\eta}^2_{\text{Within-Pair}}$	$\hat{\eta}^2_{\text{Sample Size}}$
F	.95	.02	< .01 (not sig.)
PSAR	.47	not sig.	.20 (not sig.)
MHCHICI	.29	.05	.63 (.09)
MHFCI	.83	.04	not sig. (not sig.)
ART	.28	not sig.	.49 (.02)
SHARPCHI	.46	.04	.40 (not sig.)
SHARPF	.08	.01	.88 (not sig.)

F = parametric F-test, PSAR = Puri and Sen aligned-rank test, MHCHICI = McKean-Hettmansperger chi-square test with confidence interval estimate of τ, MHFCI = McKean-Hettmansperger F test with confidence interval estimate, ART=Aligned-rank transform, SHARPCHI=Serlin-Harwell modified aligned-rank chi-square test, SHARPF = Serlin-Harwell modified aligned-rank F-test. Values in parentheses for $\hat{\eta}^2_{\text{Sample Size}}$ are based on restricting the sample size to N = 60 or 80.

Table 5

Estimated Power Values by Distribution and Sample Size

N	Distribution	F	PSAR	MHCHICI	MHFCI	ART	SHARPCHI	SHARPF
20	Normal	.6984	.3971	.7405&	.6543	.2432	.6145	.4423
20	Moderately skewed/kurtic	.7002&	.3871	.7416&	.6581	.2325	.6240	.4559
20	Skewed/kurtic	.7024&	.3738	.7407&	.6615	.2272	.6314	.4674
20	Cauchy	.7430&	.2757	.7751&	.7100&	.1439	.7248	.5690
60	Normal	.6909	.4816	.6848&	.6598	.4452	.6573	.6281
60	Moderately skewed/kurtic	.6919&	.4546	.6785&	.6528	.4201	.6668	.6365
60	Skewed/kurtic	.6865&	.4329	.6676&	.6457	.3997&	.6692&	.6418
60	Cauchy	.6750&	.2813&	.6462&	.6281&	.2545&	.7059&	.6819&

& = estimated Type I error rate > .0561. F = parametric F-test, PSAR = Puri and Sen aligned-rank test, MHCHICI = McKean-Hettmansperger chi-square test with confidence interval estimate of τ , MHFCI = McKean-Hettmansperger F test with confidence interval estimate, ART=Aligned-rank transform, SHARPCHI=Serlin-Harwell modified aligned-rank chi-square test, SHARPF = Serlin-Harwell modified aligned-rank F-test. Each tabled value is the proportion of rejections across 20,000 replications.

Appendix A

Estimating τ

In this section we briefly describe the τ scale parameter, its use in the McKean-Hettmansperger tests, and how it was estimated in our simulation. Initially it was assumed that the residuals had density f(ε) and that τ was the scale parameter of this density. Unfortunately, τ cannot typically be estimated in a simple fashion because f(ε) is unknown.

The essence of the method using the standardized length of a 90% confidence interval to estimate is τ is to create a family of confidence intervals about a point in the center of the density. These confidence levels depend only on the large-sample Wilcoxon signed-rank null distribution and not on the underlying distribution of the residuals. The lengths of these confidence intervals are then used to estimate this parameter, say τ_{CI}. The choice of 1- α is crucial in estimating τ_{CI}. Hettmansperger and McKean (1983) reported that .90 was a suitable choice, McKean and Sievers (1989) suggested that .98 was a suitable choice, and McKean and Sheather (1991) indicated that larger 1- α values are needed when the ratio of sample size to the number of parameters is less than 5; for ratios greater than 5, a 1- α of .80 appears to be sufficient. RREGRESS uses .90 by default.

Following Aubuchon and Hettmansperger (1984), $\hat{\tau}_{CI} = N^{1/2}(L_{.95} - L_{.05})/(2^* \; t_{1-\alpha})$ where L_α is a cutpoint (described below) and t is a critical t-value. First, obtain N residuals after fitting a full model and then compute the N(N+1)/2 pairwise (Walsh) averages among the residuals. For the Wilcoxon signed-rank statistic T, $\mu_T = N(N+1)/4$ and

$\sigma^2_T = N(N+1)(2N+1)/24$. Next, find $C_{.05} = \mu_T - 1.645\,\sigma_T$, rounded down to the next lower integer. After sorting the Walsh averages the $(C_{.05}+1)$st = $L_{.05}$ and $(N(N+1)/2 - C_{.05})$th = $L_{.95}$ values are the confidence interval limits, and the difference between the limits is the estimated confidence interval length, say $\hat{\delta}$. Hettmansperger and McKean (1977) suggested an estimator of the form:

$$\hat{\tau}_{CI} = [(N\,\hat{\delta}/2)]/[(N-q-1)^{1/2} t_{.95}] \qquad \text{(A1)}$$

where $t_{.95}$ is a t-value corresponding to the 95th percentile of a t-distribution with N-q-1 degrees of freedom and $[N/(N-q-1)]^{1/2}$ is an adjustment factor. The rationale for the adjustment is that the statistic in equation (A1) is biased because the residuals are correlated, shrinking their variance. A drawback of $\hat{\tau}_{CI}$ is that this estimator is only consistent if the residual distribution is symmetric.

A second method to estimate τ uses the kernel density estimation method described in Tapia and Thompson (1978), Maritz (1995), and Hettmansperger (1984). Once again τ is estimated (say, $\hat{\tau}_{WIN}$) from the density $f(\varepsilon)$. The asymptotic formula for τ_{WIN} is

$$\tau_{WIN} = 1/[(12)^{1/2} \int f^2(\varepsilon)\, d\varepsilon \qquad \text{(A2)}$$

$$= 1/[(12)^{1/2} \int f(\varepsilon) dF(\varepsilon)$$

where $\int f^2(\varepsilon)\, d\varepsilon$ = mean density and $F(\varepsilon)$ is a cumulative distribution function (cdf). Kernel estimation essentially uses a histogram to estimate the density and requires that we use the data twice, once to estimate $f(\varepsilon_i)$, say, $\hat{f}(\varepsilon_i)$, and once to estimate an empirical cdf, say $\hat{F}(\varepsilon_i)$.

Based on the formula $\int f(\varepsilon)dF(\varepsilon)$ in equation (A2), Hettmansperger (1984) proposed replacing $f(\varepsilon)$ by a kernel estimate and estimating $F(\varepsilon)$ with $\hat{F}(\varepsilon_i)$. Letting $\hat{\tau} = 1/[(12)^{1/2}\, \hat{\gamma}^*]$ ($(12)^{1/2}$ comes from McKean and Hettmansperger's use of Wilcoxon scores), define

$$\hat{\gamma}^* = \int \hat{f}(\varepsilon)\, d\hat{F}(\varepsilon) \qquad \text{(A3)}$$

The density $\hat{f}(\varepsilon_i)$ is estimated as

$$\hat{f}(\varepsilon_i) = N^{-2} h_N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} w[(r_i - r_j)/h_N],\ i \neq j \qquad \text{(A4)}$$

where w is a density that is symmetric about 0, w( ) is a window, and h_N is the window width. This method requires that the window function and the window width be selected. Many authors have pointed out that the choice of window function has little impact on the results (e.g., Bean & Tsokos, 1980), and Minitab's RREGRESS command uses a uniform density. The choice of the window width h_N, on the other hand, is crucial, and is similar to the choice of the confidence level in the confidence interval method. Large window widths lead to density estimates with small variance but substantial bias, whereas small window widths lead to density estimates with larger variance but smaller bias (Maritz, 1995, p. 28). Several authors have argued that minimizing bias is the more important of the two (e.g., Aubuchon & Hettmansperger, 1984; Bean &

Tsokos, 1980).

Aubuchon and Hettmansperger (1984) modified equation (A4) to reduce bias, producing

$$\hat{\gamma}^* = 1/[N\hat{K}] + 1/[(N(N-1)\,h_N]\sum_{i \neq j} w[(r_i - r_j)/h_N] \tag{A5}$$

where $\hat{K}$=4.1078$\hat{\delta}$ ($\hat{\delta}$ = sample interquartile range computed on the full model residuals), $h_N = \hat{K}N^{-1/2}$, and

$\sum_{i \neq j} w[(r_i - r_j)/h_N]$ requires that we compute the N(N-1) possible pairwise differences among the full model residuals and assign 1 if a pairwise difference divided by h_N is between -.50 and -.50, and 0 otherwise. (Aubuchon and Hettmansperger (1984) showed that using equation (A5) rather than (A4) reduces the bias in estimating τ from $O(N^{-2/3})$ to $O(N^{-1})$). The use of equation (A5) requires specifying the scale of the underlying distribution as well as its shape. RREGRESS assumes that the distribution is normal, which is where 4.1078 comes from (see Hettmansperger, 1984, p. 249). Finally,

$$\hat{\tau}_{INT} = 1/[(12)^{1/2}\hat{\gamma}^*] \tag{A6}$$

An advantage of this method is that estimates of τ_{WIN} are consistent without assuming symmetry of the underlying distribution. However, the statistic in equation (A6) is still biased because the residuals are correlated. Various corrections have been proposed, and the RREGRESS command uses $[N/(N-q-1)]^{1/2}$. The confidence interval (CI) and window (WIN) estimators are (asymptotically) equally accurate but show differences in small sample performance (Draper, 1988). We computed both in our simulation study for each McKean and Hettmansperger test.

Appendix B

Generating Data To Satisfy the Type I Error and Power Conditions

<u>Ho True</u>

The calculations were based on inverting the general correlation matrix

$$\underline{\rho} = \begin{array}{c|ccccc} & y & x_1 & x_2 & x_3 & x_4 \\ \hline y & 1 & \rho_{yx1} & \rho_{yx2} & \rho_{yx3} & \rho_{yx4} \\ x_1 & & 1 & \rho_{x1x2} & \rho_{x1x3} & \rho_{x1x4} \\ x_2 & & & 1 & \rho_{x2x3} & \rho_{x2x3} \\ x_3 & & & & 1 & \rho_{x3x4} \\ x_4 & & & & & 1 \end{array}$$

Here correlations with y appear in the first row and column, correlations for the reduced set of predictors (x_1-x_2) in rows and columns 2 and 3, and correlations for the x_3-x_4 predictors in rows and columns 4 and 5. Inverting this matrix results in the element in the first row and column, $R^{-1}_{11}=1/(1-R^2)$ in predicting y from the other variables in the matrix. We began with the reduced model matrix with $\rho_{x1x2} = \rho_{x3x4} = 0$,

$$\underline{\rho}_{reduced} = \begin{bmatrix} 1 & \rho_{yx1} & \rho_{yx2} \\ \rho_{yx1} & 1 & 0 \\ \rho_{yx2} & 0 & 1 \end{bmatrix}$$

and the element R^{-1}_{11} is $1/(1-2\rho^2_{yx1})$. Solving for ρ^2_{yx1} yields $\rho^2_{yx1}=.1$ and $\rho_{yx1}=(.1)^{1/2}$ (we drop the subscripts on ρ involving x from here on since, for example, $\rho_{yx1} = \rho_{yx2}$). Thus, the correlation between y and x_1 and between y and x_2 when the correlation between x_1 and x_2 (and between x_3 and x_4) equals zero was $(.1)^{1/2}=.3162$. Repeating this process for

$\rho_{x1x2} = \rho_{x3x4} = .3$, the element R^{-1}_{11} is $.91/(.91-1.4\rho^2_{yx1})$. Solving, we get $\rho_{yx1} = (.13)^{1/2}$, so the correlation between y and x_1 and between y and x_2 when the correlation between x_1 and x_2 (and between x_3 and x_4) equals 0.3 was $(.13)^{1/2}$=.3605.

When the correlation between x_1 and x_2 (and between x_3 and x_4) equals zero and the reduced model accounts for R^2=.2, under the null hypothesis, the full model must also account for R^2=2. For the matrix (assuming $\rho_{x1x2} = \rho_{x3x4} = 0$),

$$
\begin{array}{cc}
 & \begin{array}{ccccc} y & x_1 & x_2 & x_3 & x_4 \end{array} \\
\begin{array}{c} y \\ x1 \\ x2 \\ x3 \\ x4 \end{array} &
\begin{pmatrix}
1 & .3162 & .3162 & \rho_{yx} & \rho_{yx} \\
.3162 & 1 & 0 & .3 & .3 \\
.3162 & 0 & 1 & .3 & .3 \\
\rho_{yx}\ .3 & .3 & 1 & 0 & \\
\rho_{yx}\ .3 & .3 & 0 & 1 &
\end{pmatrix}
\end{array}
$$

the element R^{-1}_{11} is $\dfrac{.64}{.8(.64)-2(\rho_{YX}-.6\sqrt{.1})^2}$, and so $R^2 = .2 + \dfrac{2}{.64}(\rho_{YX}-.6\sqrt{.1})^2$. For the case when Ho is true, set $\rho_{yx} = (.6)(.3162)$. Repeating this for the case for $\rho_{x1x2} = \rho_{x3x4} = .30$ produces the element R^{-1}_{11} =

$$.91\left[\frac{.637(.931)}{.728(.637)(.931)-1.27(.91\rho_{YX}-.42\sqrt{.13})^2}\right] \text{ and so}$$

$$R^2 = .2 + \frac{1.27}{.91(.637)(.931)}(.91\rho_{YX}-.42\sqrt{.13})^2.$$ For Ho true set $\rho_{YX} = \dfrac{.42}{.91}\sqrt{.13} = .1664$.

Ho False

Assuming the x_{ik} are fixed, the above results can be used to calculate the correlations needed to generate a power of 0.7. In each of the R^2 formulae, we see that part of the change in R^2 depends on the correlation between y and x_3 and between y and x_4. What is needed is to compute the change in R-square (over the reduced model R-square of 0.2) needed to yield a power of 0.7 (The noncentrality parameter here is $N\Delta R^2/(1-R^2)$). The NCSSCALC program (downloadable at http://www.ncss.com/download.html) was used to calculate the required change in R^2 (just to make sure the results were replicated using the G-Power program downloadable at http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/). The above formulas were then used to solve for ρ_{YX}. For $\rho_{x1x2} = \rho_{x3x4} = 0$ and N=20 the resulting correlation matrix was:

$$\begin{pmatrix} 1 & \sqrt{.1} & \sqrt{.1} & .476386 & .476386 \\ \sqrt{.1} & 1 & 0 & 0.3 & 0.3 \\ \sqrt{.1} & 0 & 1 & 0.3 & 0.3 \\ .476386 & 0.3 & 0.3 & 1 & 0 \\ .476386 & 0.3 & 0.3 & 0 & 1 \end{pmatrix}$$

However, an adjustment of ρ_{YX} was needed because the x_{ik} were random, meaning that the F follows a confluent hypergeometric distribution. The result of the adjustment was that $\rho_{YX} = .476386$ was changed to .497 to generate the desired power of 0.7 under these conditions. For $\rho_{x1x2} = \rho_{x3x4} = .30$ and N=20, the resulting correlation matrix was:

$$\begin{pmatrix} 1 & \sqrt{.13} & \sqrt{.13} & .529404 & .529404 \\ \sqrt{.13} & 1 & 0.3 & 0.3 & 0.3 \\ \sqrt{.13} & 0.3 & 1 & 0.3 & 0.3 \\ .529404 & 0.3 & 0.3 & 1 & 0.3 \\ .529404 & 0.3 & 0.3 & 0.3 & 1 \end{pmatrix}$$

The adjustment changed $\rho_{YX} = .529404$ to .559. For $\rho_{x1x2} = \rho_{x3x4} = .30$ and N=60, the resulting correlation matrix was:

$$\begin{pmatrix} 1 & \sqrt{.13} & \sqrt{.13} & .387814 & .387814 \\ \sqrt{.13} & 1 & 0.3 & 0.3 & 0.3 \\ \sqrt{.13} & 0.3 & 1 & 0.3 & 0.3 \\ .387814 & 0.3 & 0.3 & 1 & 0.3 \\ .387814 & 0.3 & 0.3 & 0.3 & 1 \end{pmatrix}$$

The adjustment changed $\rho_{YX} = .387814$ to .389.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM033726

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: An Empirical Study of Eight Nonparametric Tests in Hierarchical Regression

Author(s): Michael R. Harwell, Ronald C. Serlin

Corporate Source:	Publication Date:
Ø	

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, ANO IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☑	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Michael R. Harwell	Printed Name/Position/Title: Professor	
Organization/Address: Univ. of Minnesota, 323 Burton Hall, 178 Pillsbury Dr. SE. Minneapolis, MN 55455	Telephone: 612-625-0196	FAX:
	E-Mail Address: harwe001@umn.edu	Date: 3/13/02

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)